

38



(11) Publication number:

0 079 739
A2

(12)

EUROPEAN PATENT APPLICATION

(21) Application number: B2305926.6

(51) Int. Cl.³: C 12 N 15/00
C 12 N 1/00, C 12 P 21/02
C 07 H 21/04, C 07 C 103/52
//C12R1/19, C12R1/865

(22) Date of filing: 08.11.82

(30) Priority: 12.11.81 US 320632

(71) Applicant: THE UPJOHN COMPANY
301 Henrietta Street
Kalamazoo, Michigan 49001(US)

(43) Date of publication of application:
25.05.83 Bulletin 83/21

(72) Inventor: Dugaliczyk, Achilles
c/o The Upjohn Company 301 Henrietta Street
Kalamazoo Michigan 49001(US)

(84) Designated Contracting States:
BE CH DE FR GB IT LI NL SE

(74) Representative: Perry, Robert Edward et al,
GILL JENNINGS & EVERY 53-64 Chancery Lane
London WC2A 1HN(GB)

(54) Albumin-based nucleotides, their replication and use, and plasmids for use therein.

(57) The DNA sequence coding for human serum albumin has been isolated and inserted as two fragments into two novel plasmids which can be replicated in *E. coli*. These novel fragments can be joined to provide a unitary DNA sequence which then can be cloned into a suitable host, e.g. *E. coli*, for the expression of human serum albumin (which is used extensively in medical practice in treating shock conditions).

EP 0 079 739 A2

ALBUMIN-BASED NUCLEOTIDES, THEIR REPLICATION
AND USE, AND PLASMIDS FOR USE THEREIN

This invention relates to nucleotides related to human serum albumin (HSA), their replication and use, and plasmids (and host substances) for use therein.

The gene for serum albumin is regulated in development. On the other hand, serum albumin is synthesised in mammals by the adult liver, and its plateau in adulthood. The embryonic liver and yolk sac, on the other hand, produce predominantly α -fetoprotein, but the synthesis decreases drastically after birth. Recently, Law et al determined the complete sequence of mouse α -fetoprotein mRNA, Nature 291 (1981) 201-205. The structure revealed extensive homology to mammalian serum albumin, indicating that the two proteins are encoded in the same gene family. Similar conclusions have been reached from studies on the α -fetoprotein genes of the rat and the mouse; see Jagodzinski et al, Proc. Natl. Acad. Sci. USA, 78 (1981) 3521-3525, and Gorin et al, J. Biol. Chem. 256 (1981) 1954-1959.

The complete nucleotide sequence of human serum mRNA has been determined from recombinant cDNA clones and from a primer-extended cDNA synthesis on the mRNA template. The sequence comprises 2,078 nucleotides, starting upstream of a potential ribosome binding site in the 5'-untranslated region. It contains all the translated codons and extends into the poly(A) at the 3'-terminus. Part of the translated sequence codes for a hydrophobic prepeptide met-lys-trp-val-thr-phe-ile-ser-leu-leu-phe-leu-phe-ser-ser-ala-tyr-ser, followed by a basic propeptide arg-gly-val-phe-arg-arg. These signal peptides are absent from mature serum albumin and, so far, have not been identified in their nascent state in humans. A remaining 1,755 nucleotides of the translated mRNA sequence code for 585 amino acids which are in agreement, with few exceptions, with the published amino acid data for human serum albumin. The mRNA sequence verifies and refines the repeating homology in the triple-domain structure of the serum albumin molecule.

DETAILED DESCRIPTION OF THE INVENTION

Human serum albumin cDNA is cloned into the PstI site of plasmid pBR322 by the oligo(dG)-oligo(dC) tailing technique. Plasmid DNA was isolated from 97 positive colonies which hybridized to the enriched 5 albumin cDNA probe, and the recombinant plasmid pH A36 was found to contain the largest insert of an albumin cDNA sequence. Its restriction endonuclease map is shown in the drawing, together with a restriction map of the primer-extended plasmid clone pH A206. The latter was obtained in a second transformation experiment after initiating 10 the cDNA synthesis from an internal primer. This primer was a 91 base pairs long DNA fragment, MspI(152)-TaqI(182/3), isolated from pH A36. The two plasmids, pH A36 and pH A206, share 0.15 kb of homologous DNA. Together, they encode the entire sequence for human serum albumin, starting with the CTT codon for leu -10 of the prepeptide and extending 15 into the 3'-untranslated region of poly(A).

Sequence of the Albumin cDNA. The sequence was determined for the most part on both DNA strands to ensure accuracy. All of the restriction sites used to end-label DNA fragments were sequenced across by 20 labeling a neighboring restriction site. The entire nucleotide sequence of the serum albumin mRNA, as determined from the cloned DNA in pH A36, pH A206, and from the primer-extended cDNA at the 5'-terminus of the message, is shown in the following Table 1. The inferred amino acid sequence is also indicated. The mRNA length is 2,078 nucleotides, of which 38 represent the 5'-untranslated region, 54 identify a 25 prepeptide of 18 amino acids, 18 identify a propeptide of 6 amino acids, 1,755 code for the known 585 amino acids of serum albumin, 189 make up the 3'-untranslated region and 24 are the poly(A) sequence. Nucleotides 5 to 15 (-34 to -24) in the 5'-untranslated region (Table 30 1) are complementary to a 3'-terminal region of eukaryotic 18S RNA [Azad, A.A. and Deacon, N.J. (1980) Nucl. Acids Res. 8, 4365-4376] and thus could represent a ribosome binding site:

(5')...T T^C T C T T C T G T.....albumin mRNA
35 (3')...G A G G A A G G G C G U C C m₂⁶A m₂⁶A.....18S RNA

The translated portion of the mRNA sequence codes for the signal peptide and the main body of the albumin polypeptide chain. The

signal peptide is composed of a hydrophobic prepeptide of 18 amino acids and a basic propeptide of 6 amino acids (Table 1). Since pre-peptides are removed from nascent secretory proteins (like albumin) in the endoplasmic reticulum, they are seen only in vitro in heterologous 5 translation systems. As yet, they have not been found within cells [Judah, J.D. and Quinn, P.S. (1977) FEBS 11th Mtg., Copenhagen 50, 21-29; and Strauss, A.W., Donohue, A.M., Bennett, C.D., Rodkey, J.A. and Alberts, A.W. (1977) Proc. Natl. Acad. Sci. USA 74, 1358-1362]. This is the first report of the presence and the sequence of a pre- 10 peptide for human serum albumin. As it is with other secretory proteins, the conversion of proalbumin to albumin takes place in the Golgi vesicles, and the enzyme responsible for this cleavage is probably cathepsin B [Judah, J.D. and Quinn, P.S. (1978) Nature 271, 384-385]. This is also a first report on the sequence of the pro- 15 peptide for normal human serum albumin.

At the 3'-end of the message, the putative polyadenylation signal sequence, AATAAA, is located 164 nucleotides downstream from the amino acid termination codon TAA and 16 nucleotides upstream from the beginning of the poly(A) sequence. Another characteristic sequence 20 located near the polyadenylation site has been identified by Benoist, et al. [Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142]; the consensus sequence from several mRNAs was concluded as TTTTCACTGC. A similar sequence, TTTTCTCTGT, is located 19 nucleotides upstream from the AATAAA hexanucleotide in the 25 human albumin mRNA (Table 1).

0079739

4083

-4-

TABLE 1

	35	25	20	15	10	5
			-18 p r o		-10	
			Met lys trp val tlu phe ile ser leu leu phe leu phe ser			
			GCT TTT TCT CTC TGT GCA ACC CCC AAC GCC CTT TGG CAC A ATG AAG TGG GTC ACC TTT ATT TCC CTT CTT AGT AGC (30)			
-1	-6	p r o	-1	1		
ser ala tyr ser arg gly val phe arg asp ala his lys ser glu val ala his arg phe lys asp leu gly glu asn phe lys						
TCC CCT TAT TCC ACC GCT GTC TTT CGT CAT GCA CAC AAC AGT GAC CCT CCT CGC ATT GCC ATT GCT GTC TAT CCT CAG TAT						
CCC TTG GTC TCG ATC ATT GCC ATT GCT GTC TCA GCT GAA ATT TGT GAC AAA TCA CTT CAT ACC CCT ATT CGA GAT CAT GTC ATT GCA ATT (170)						
21	ala leu val leu ile ala phe ala gln tyr leu gln gln cys pro phe glu asp his val lys leu val asn glu val thr glu phe ala					
30	lys thr cys val ala asp glu ser ala glu asn cys asp lys ser leu his thr leu phe gly asp lys leu cys thr val ala thr leu					
51 53	90 91	60 62	70 75	100 101	110	80
lys thr cys val ala asp glu ser ala asp cys cys ala lys gln gln pro gly arg asn glu cys phe leu gln his lys asp asp asn pro						
AAA ACA TGT GTC ATT GCT GAT GAC TCA GCT GAA ATT GCA ATT GCA CCT GGC AGA ATT GAA TGC TTC CAA CAC AAA GAT GAC AAC CCA (440)						
81	CGT GAA ACC TAT CGT GAA ATG GCT GAC TGC TGT GCA ATT CCT GCA ATT (420)					
111	arg glu thr tyr gly glu met ala asp val met cys thr ala phe his asp asn glu glu thr phe leu lys tyr leu try					
asn leu pro arg leu val arg pro glu val asp val met cys thr ala phe his arg tyr lys ala ala phe thr glu cys cys gln						
AAC CTC CCC CGA TTG GTC AGA CCA GAC GTT CAT GTC ATG TCC ACT GCT TTT CAT GAC ATT GAA TGC ATT AAA AGC TAT AAA GCT TAT TTA TAT (330)						
141	150	120 124	130	160	168 169 170	
glu ile ala arg arg his pro tyr phe tyr ala pro glu leu leu phe ala lys arg tyr lys ala ala phe thr glu cys cys gln						
GAA ATT GCC AGA AGA CAT CCT TAC TTT TAT GGC CCC GAA CTC CTT TTC ATT CCT ATT AAA AGC TAT AAA GCT TAT TTA GAA TGC TCA ATT (620)						
171	177	180	190			
ala ala asp lys ala ala cys leu leu pro lys leu asp glu leu arg asp glu ala lys ala ser ser ala lys gln arg leu lys cys						
CCT CCT GAT AAA CCT GCT GCC TGC CTC CCA AAC CTC GAT GAA ATT CCT GCT GTC ATT GCA ATT CCT GCT GTC ATT GCA ATT (710)						
211	210	220	230			
ala ser leu gln lys phe gly glu arg ala phe lys ala arg leu ser gln arg phe pro lys ala glu phe ala glu						
CCC AGT CTC CCA AAA TTT GCA GAA ACA CCT TTC AAA GCA TGC GCA CCT AGC CTC AGC CAG AGA ATT CCC AAA CCT GCA GAA (360)						

0079739

-5-

4083

35 20 15 10 5
231 240 245 246 250 253 260
val ser lys leu val thr asp leu thr lys val his thr glu cys oys his gyl asp leu leu glu cys ala asp arg ala asp leu
GTT TCC AAG TTA GTC ACA GAT CTT ACC AAA GTC CAC ACC GAA TGC TCC CAT GCA GAT CTC CCT GAA TGT GAT GAC AGG CGG GAC CTT (890)
261 265 270 278 279 280 289 290
ala lys tyr lle cys glu asn gln asp ser lle ser ser lys leu lys glu cys cys glu lys pro leu leu glu lys ser his cys lle
GCC AAG TAT ATC TGT GAA AAT CAA GAT TCG ATC TCC AGT AAA CTG AAC GAA TGC TGT GAA AAA CCT CTC TTG GAA AAA TGT CAT ATT (980)
291 300 321 330 340 350
ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala
GCC GAA GTC GAA AAT GAT GAG ATG CCT GCT GAC TTG CCT TCA TTA GET CCT GAA ACT AAC CAT GTT TGT AAC TAT GTC ATT (1070)
321 330 340 350
glu ala lys asp val phe leu tyr glu met phe leu tyr ala arg his pro asp tyr ser val leu leu arg leu ala
GAC GCA AAC GAT GTC TTC TTG GCC ATG TTT TTG TAT GAA TAT GCA AGA AGG CAT CCT GAT TAC TAC TCT GTC GTC CTC AGA CTT GCA (1160)
351 360 361 369 370 380
lys thr tyr glu thr thr leu glu lys cys cys ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu
AAG ACA TAT GAA ACC ACT CTA GAG AAC TGC TGT GCC CCT GCA GAT CCT CAT GAA TGC TAT GCA AAA GTG TTC GAT GAA TTT AAA CCT CCTT (1250)
381 390 392 400 410
val glu glu pro gln esn leu lle lys gln asn cys glu leu phe glu gln leu gln Tyr lys phe gln asn ala leu leu val arg
GTC GAA GAG CCT CAG AAC TAA TTA ATC AAA CAA CCT CTC TCA ACT CCT GTA GAC GTC TCA AGA AAC CTA GCA AAA GTG GGC AGC AAA TGT TGT AAA CAT (1340)
420 430 440 450 460 461 470
tyr thr lys val pro gln val ser thr pro thr leu val glu val val ser arg asn leu gln lys val gln lys thr pro val ser
TAC ACC AAC AAA GTA CCC CAA GTC TCA ACT CCT GTA GAC GTC TCA AGA AAC CTC TCA TGC TGT CCT GAT GAA ACA TAC GCA GTC ACT (1430)
441 448 450 460 461 470
pro glu ala lys arg met pro oys ala glu asp tyr leu ser val gln leu gln lys thr tyr val pro lys
CCT GCA AAA AGA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTC GTC AAC CAG CCA CCA TGC TTT TCA CCT CTC GAA GTC CAT GAA ACA TAC GCA GTC ACT (1520)
471 476 477 480 490 500
asp arg val thr lys cys oys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp glu thr tyr val pro lys
CAC AGA GTC ACC AAA TGC TCC ACA GAA TCC TGT GTC AAC CTC CCT GTC GAA GTC GCA CCA CCA TGC TTT TCA CCT CTC GAA GTC CAT GAA ACA TAC GCA GTC ACT (1610)
501 510 514 520 530
glu phe esn ala glu thr phe thr phe his ala asp lle cys thr leu ser gln lle lys lys gln thr ala leu val
GAC TTT AAT GCT GAA ACA ATC TGC ACC TTC CAT GCA GAT ATA TGC ACA CTT TCT GAC AAG GAG AGA CAA ATC AAC AAA ACT GCA CTT GTC (1700)

35		20	5
30		15	550
25		10	559 560
20		5	

531
 glu leu val lys his lys pro lys ala thr lys glu qin leu lys ala val met asp asp phe ala ala phe val glu lys cys cys lys
 GAG CTC GTG AAA CAC AAC CCC AAG GCA ACA AAA GAG CAA CTC AAA GCT GGT ATC GAT GAT TTC GCT GCT GCT GTC TGC TGR AAC (1790)

540
 ala asp lys glu thr cys phe ala glu qin glu ala lys leu val ala ser qin ala ala leu qly leu ter ter
 GCT GAC GAT AAC GAC ACC TGC TTT GCC GAG GGT AAA AAA CTT GCT GCT GCA GCT CAA GCT GCA GCT GCA GCT GCA GCT GCA GCT GCA (1883)

561
 ter ter
 CATCTAGCCTACCATGAGATAAGACAAGAAATGAACATCAAAAGCTTATTCACTGGTTCTTCTTCTTCTTCTTCTTCTTGTGTGTTAAACCCCAACACCTGTCTAAAAAACATAAATTTCCTTAA (2002)

567
 TCATTTGGCTCTTTCTCTGTCTTCAATTAAATAAAATGAAAGAATCTAA..... 20AA (2078)

570
 GCT

580
 ter

Following are examples which illustrate procedures, including the best mode, for practicing the invention. These examples should not be construed as limiting. All percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted.

5 Example 1 Isolation of Messenger RNA

Human liver mRNA was obtained following the procedure of Chirgwin, et al [Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299]. Immunoprecipitation of albumin containing polysomes was performed according to Taylor and 10 Tse [Taylor, J.M. and Tse, T.P.H. (1976) J. Biol. Chem. 251, 7461-7467]. In vitro translation of mRNA was carried out in a reticulocyte cell-free system, following the instruction of the manufacturer (New England Nuclear). The translation products were separated electrophoretically according to Laemmli [Laemmli, J.K. (1970) Nature 227, 15 680-685.

15 Example 2 Cloning Procedures

Double stranded cDNA was synthesized as described previously [Law, S., Tamaoki, T., Kreuzaler, F. and Dugaiczyk, A. (1980) Gene 10, 53-61]. It was annealed to PstI-linearized pBR322 DNA [Bolivar, F., 20 Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crossa, J.H. and Falkow, S. (1977) Gene 2, 95-113] that had been tailed with 15 dG residues/3'-terminus [Dugaiczyk, A., Robberson, D.L. and Ullrich, A. (1980) Biochemistry 19, 5869-5873]. The annealed DNA was used to transform E. coli strain RR1, as detailed previously [Law, 25 S., et al., Ibid.]. The albumin clones were selected using the colony hybridization method of Grunstein and Hogness [Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965], with [³²P]-labeled cDNA synthesized with the immunoprecipitated polysomal mRNA as template.

30 As shown in Example 5, plasmids pH A36 and pH A206 were deposited in E. coli HB101 hosts. The plasmids were obtained from E. coli RR1 hosts, described in this example, and transformed into E. coli HR101 by standard procedures well known to those of ordinary skill in this art. The E. coli RR1 hosts were lysed and then centrifuged to 35 separate the chromosomal DNA, cell DNA and plasmid DNA. The plasmid DNA, remaining in the supernatant, is precipitated with ethanol and the precipitate is resuspended in buffer, e.g., TCM (10mM Tris-HCl, pH 8.0, 10 mM CaCl₂, 10 mM MgCl₂). The cells for transformation are

prepared as follows: 120 ml of L-broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl) are inoculated with an 18 hour culture of HB101 NRRL B-11371 and grown to an optical density of 0.6 at 600 nm. Cells are washed in cold 100 mM NaCl and resuspended for 15 minutes in 20 ml 5 chilled 50 mM CaCl₂. Bacteria are then concentrated to one-tenth of this volume in CaCl₂ and mixed 2:1 (v:v) with annealed plasmid DNA, prepared as described above. After chilling the cell-DNA mixture for 15 minutes, it is heat shocked at 42°C for 2 minutes, then allowed to equilibrate at room temperature for ten minutes before addition of 10 L-broth 10 times the volume of the cell-DNA suspension. Transformed cells are incubated in broth at 37°C for one hour before inoculating selective media (L-agar plus 10 µg/ml tetracycline) with 200 µl/plate. Plates are incubated at 37°C for 48 hours to allow the growth of transformants.

15 Example 3 Mapping of Restriction Endonuclease Sites

Restriction endonucleases were obtained from Bethesda Research Laboratories and New England Biolabs and were used according to the manufacturers' instructions. The digested DNA fragments were analyzed electrophoretically on agarose [Helling, R.B., Goodman, H.M. and 20 Boyer, H.W. (1974) J. Virol. 14, 1235-1244] or acrylamide [Dingman, C., Fisher, M.P. and Kakefuda, T. (1972) Biochemistry 11, 1242-1250] gels.

Example 4 DNA Sequencing

DNA fragments were dephosphorylated with bacterial alkaline 25 phosphatase (Worthington) and labeled at the 5'-ends with poly-nucleotide kinase (Boehringer-Mannheim) and γ [³²P]ATP. Following digestion with a second restriction endonuclease and electrophoretic separation of the fragments, DNA sequence determination was done according to the procedure of Maxam and Gilbert [Maxam, A. and 30 Gilbert, W. (1980) Methods Enzym. 65, 499-560] and the degradation products were separated electrophoretically on 0.4 mm acrylamide gels as described by Sanger and Coulson [Sanger, F. and Coulson, R. (1978) FEBS Letters 87, 107-110].

Example 5 Recombinant Plasmids pH A36 and pH A206

35 As disclosed in Example 2, albumin clones were selected by hybridizing to the enriched albumin cDNA probe. Plasmid pH A36 contained the largest insert of an albumin cDNA sequence. Both plasmids pH A36 and pH A206 have been deposited in a viable E. coli host in the

permanent collection of the Northern Regional Research Laboratory (NRRL), U.S. Department of Agriculture, Peoria, Illinois, U.S.A. Their accession numbers in this repository are as follows:

HB101(pHA36) - NRRL B-12551

5 HB101(pHA206) - NRRL B-12550

E. coli HB101 is a known and widely available host microbe. Its NRRL accession number is NRRL B-11371.

10 NRRL B-12550 and NRRL B-12551 are available to the public. ~~upon the grant of a patent. It should be understood that the availability of these deposits does not constitute a license to practice the subject invention in derogation of patent rights granted with the subject instrument by governmental action.~~

15 E. coli RR1 and E. coli HB101 are known and widely available host microbes. Their NRRL accession numbers are NRRL B-12186 and NRRL B-11371, respectively.

pBR322 is a well known and widely available plasmid. It can be obtained from the following host deposit by standard procedures:

NRRL B-12014 - E. coli RR1 (pBR322).

20 YEpl6 is a well known and widely available yeast episomal plasmid. It can be obtained from the following host deposit by standard procedures:

E. coli HB101 (YEpl6) - NRRL B-12093.

Example 6 Assembly of the Serum Albumin Gene

25 Assembling the pieces together is a straightforward task of restriction enzymology. There is only one MspI site in the overlapping DNA sequence of the two cDNA clones. Two enzymatic steps of (i) MspI digestion of the two DNAs, followed by (ii) the use of ligase, an enzyme that seals DNA fragments, will give the desired product. Although two other undesired DNA species will also be obtained in the 30 course of this recombination reaction, both of them will differ substantially in size. Thus, separation and isolation of the desired DNA species will be achieved.

The assembled DNA clone can be used to transform two types of cells:

35 (a) Escherichia coli

(b) Saccharomyces cerevisiae

(a) The vector of choice is plasmid pBR322, the same that has

been successfully used for cloning of the two fragmented pieces of the serum albumin cDNA.

(b) In order to transform yeast with the serum albumin structural gene sequence, the DNA must be inserted into one of the 5 existing yeast plasmid vectors. This can be accomplished by taking advantage of the fact that several restriction endonuclease recognition sequences are absent from the cloned serum albumin DNA. Synthetic EcoR1 DNA linkers can be ligated to the DNA fragment containing the serum albumin sequence followed by insertion (ligation) into one 10 of the yeast plasmid vectors, e.g., YEp6, at the EcoR1 cloning site. The fused chimeric plasmid can be used to transform yeast according to an established procedure [Hinnen, A., Hicks, J.B. and Fink, G.R. (1978) Proc. Natl. Acad. Sci. USA, 75, 1929]. YEp6 can be obtained from the NRRL repository, as disclosed *supra*.

15 Example 7 Expression of the Serum Albumin Gene

The main body of the structural gene will be transcribed by the E. coli or yeast enzymes. If little or no albumin is produced with the selected host, then an Escherichia coli promoter DNA sequence carrying an initiation codon, i.e., ATG, can be ligated at the beginning 20 of the serum albumin structural gene. Such elements are known and available, e.g., lac promoter used for the expression of human interferon gene in E. coli [Proc. Natl. Acad. Sci. 77, 5230 (1980)]; source of promoter DNA [Proc. Natl. Acad. Sci. 76, 760 (1979)]. Also, see Nature, Vol. 281, October 18, 1979. It has already been 25 documented that such Escherichia coli promoter sequences function well in the expression of foreign genes in Escherichia coli [Mercereau-Pujalon, O., Royal, A., Cami, B., Garapin, A., Krust, A., Gannon, I. and Kourilsky, P. (1978) Nature 275, 505; and Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Grea, R., Hirose, 30 T., Kraszewski, A., Itakura, K., and Riggs, A. (1979) Natl. Acad. Sci. USA 76, 106]. For expression in yeast, see Rose, M., Casadaban, M.J. and Botstein, D. (1981) Proc. Natl. Acad. Sci. USA 78, 2460 and 4466.

Example 8 Screening of Clones Producing Albumin

Immunological methods can be used to detect small amounts of 35 albumin made in a bacterium. Flat disks of flexible polyvinyl are coated with the IgG fraction from an immune serum and the disks are pressed onto an agar plate so that antigen released from an in situ lysed microbial colony can bind to the fixed antibody. The plastic

disk is then incubated with the same total IgG fraction labeled with radioactive iodine so that other determinants on the bound antigen can in turn bind the iodinated antibody. Radioactive areas on the disk expose X-ray film during autoradiography and thus identify colonies 5 producing the protein which is being screened for. Detailed protocols of this procedure have been published [Broome, S. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA, 75, 2746]. The purification of human serum albumin can be accomplished by using procedures well known in the art. For example, procedures disclosed in a chapter by T. 10 Peters: Purification and Properties of Serum Albumin, in: The Plasma Proteins, Putnam, Ed. Academic Press, New York, 1975, can be used.

The work described herein was all done in conformity with physical and biological containment requirements specified in the NIH Guidelines.

15

20

25

30

35

CLAIMS

1. Plasmid pH A36, having a restriction endonuclease pattern as shown in the drawing.

5

2. Plasmid pH A206, having a restriction endonuclease pattern as shown in the drawing.

3. E. coli HB101 (pHA36) having the deposit accession number
10 NRRL B-12551.

4. E. coli HB101 (pHA206) having the deposit accession number
NRRL B-12550.

15 5. A microorganism modified to contain a nucleotide sequence coding for the amino acid sequence of human serum albumin; said nucleotide sequence is as follows:

20

25

30

35

0079739

4083

-13-

ser ala tyr ser arg gly val phe arg arg asp ala his lys ser glu val ala his arg phe lys asp leu ala glu asn phe lys
TGC CCT TAT TCC ACC GGT GTC TTT CGT CAT CGG AAA GAT TTG CGA CAA AAT TTC AAA (170)

ala leu val leu ile ala phe ala gln tyr leu gln gln cys pro phe glu esp his val lys leu val asn glu val thr alu phe ala
TTC GTC TTG ATT CCC TTT GCT CAG TAT CTT CAG CAG TGT CCA AAA GAT CAT GTA AAA TTA GTC AAT CAA TTT GCA (260)

lys thr oys val ala esp glu ser ala glu asn cys esp lys ser leu his thr leu phe gly esp lys leu cys thr val ala thr leu
AAA ACA TGT GTC CCT GAT GAG TCA CCT AAA ATG TGT GCA AAA TCA CCT CAT ACC CTT TGA CAC AAA TTA TGC ACA CCT GCA ACT CTT (350)

arg glu thr tyr gly glu met ala esp cys cys ala lys gln glu pro gly arg asn glu cys ala leu aln his lys esp asn pro
CGT GAA ACC TAT CCT GAA ATG CCT GAC TCC AAC CAA CCT GGG AGA AAT GAA TGC TTC TGC CAA CAC AAA GAT GAC AAC CCA (440)

asn leu pro arg leu val esp val met cys thr ala phe his esp asn glu glu thr phe leu lys tyr leu try
AAC CTC CCC CCA TTG GTC AGA CCA GAG GTC CCT GAT GTC ATG TCC ACT GCT TTT CAT GAC AAT GAA CAG ACA TTT TGC AAA TAC TTA TAT (530)

glu ile ala arg arg his pro tyr phe tyr ala pro glu leu leu phe ala lys arg tyr lys ala phe thr glu lys cys aln
GAA ATT GCC AGA ACA CAT CCT TAC TAC TTT TAT CCC CCC GAA CTC CTC TTT GCT TTT AAC ACC TAT AAA GCT GCT TTT AGC GAA TGT TGC CAA (620)

ala ala esp lys ala ala cys leu leu pro lys leu esp glu glu lys ala ser ser ala lys aln ala leu lys cys
GCT CCT CAT AAA CCT CCC TAC CTC GAT GAA CTC AAG CTC GAT GAA CCT CGG GAT GAA GCG AGC GCT TCC TCT GCT CCC AAG AGA CTC AAC TGT (710)

ala ser leu gln lys phe gly glu arg ala phe lys ala arg leu ser gln arg phe ala ala glu glu
GCC ACT CTC CAA AAA TTT GCA GCA GCA TGC AGC CAG AGA TTT CCC AAA CCT GAG TTT GCA GAA (800)

35 30 25 20 15 10 5
 240 245 246 250 253 260
 val ser lys leu val thr asp leu thr lys val his thr glu cys his gln asp leu glu cys ala asp asp arg ala asp leu
 GTC TCC AAG TTA GTC ACA GAT CTT ACC AAA GTC CAC ACG GAA TCC TGC CAT CTC CTT GAA GAT CTC TGT GCT GAT GAC AGC GCG GAC CTT (890)
 261 265 270 278 279 280 289 290
 ala lys tyr lle cys glu een gln asp ser lle ser ser lys leu lys glu cys oys glu lys pro leu glu lys ser his cys lle
 CCC AAC TAT ATC TGT GAA AAT CAA GAT TCC ATC TCC ACT AAA CTC AAG GAA TCC TGT GAA AAA CCT CTC TGC TTG GAA AAC TAC CAC TGC ATT (980)
 291 300
 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala asp phe val glu ser lys asp val cys lys asn tyr ala
 GCA GTC GAA AAT GAT GAC ATG CCT CTC GAT TTA GCT GCT GAT TTT GTC GAA ACT AAC GAT GTT TGC AAA AAC TAT CCT GCT (1070)
 321 330
 glu ala lys asp val phe leu gly met phe leu tyr ala arg arg his pro asp tyr ser val val phe glu phe lys pro leu ala
 GAG GCA AAG GAT GTC TTC TGC TGC TGT TTT TGC TAT GAA TAT GCA AGA AGC CAT CCT CAT GAA TGC TAC TCT TGT AAC CTC AGA CCT GCC (1160)
 351 360 361
 lys thr tyr glu thr thr leu glu lys cys ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu
 AAC ACA TAT GAA ACC ACT CTA GAG AAG TCC TGT CCT CCT GCT (1250)
 381 390 392
 val glu glu pro gln asn leu lle lys gln asn cys glu leu phe glu gln leu gln lys phe gln asn ala leu leu val era
 GTC GAA GAG CCT CAG AAT TTA ATC AAA CAA AAT TGT GAG CTT TTT GAC CAG CCT GCA GAG TAC AAA TTC CAG AAC TGT TGT AAA CAT (1340)
 411 420
 tyr thr lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu gln lys val gln lys ser lys cys lys his
 TAC ACC AAG AAA GTC CCC CAA GTC TCA ACT CCA ACT CTT GTC GAG GTC TCA AGA AAC CTA GGA AAA GTC GGC AGC AAA TGT TGT AAA CAT (1430)
 441 450
 pro glu ala lys arg met pro oys ala glu asp tyr leu ser val val gln leu oys val leu his glu lys thr pro val ser
 CCT GCA GCA AAA AGA ATG CCC TGT GCA GAA GAC TAT CTA TCC TGC GTC AAC CAG TTA TGT GTC CAT GAG CAA AGC CCA GTC AGT (1520)
 471 476 477 480
 asp arg val thr lys cys thr glu ser leu val asn arg pro cys phe ser ala leu glu val asp glu thr tyr val pro lys
 GAC AGA GTC ACC AAA TGC TGC ACA GAA TCC TGT GTC AAC AGC CCA CCA CCT GTC GAA GTC GAT GAA ACA TAC GTC GCT CCC AAA (1610)
 501 510 514 520 530
 glu phe asn ala glu thr phe thr phe his ala asp lle cys thr leu ser gln lle lys lys aln thr ala leu val
 GAC TTT AAT GCT GAA ACA ACC TTC CAT GCA GAT ATA TCC ACA CTT TCT GAC AAC GAC ACA CAA ATC AAG AAA ACT GCA CCT GTC (1700)

0079739
4083

-15-

5
10
15
20
25
30
35

531 540 550 559 560
glu leu val lys his pro lys ala thr lys glu gln leu lys ala val met asp asp phe ala phe val glu lys cys cys lys
GAG CTC GTC AAA AAC CCC AAG GCA ACA AAA GAC CAA CTC AAA CCT GTT ATG GAT CAT TTC GCT CCT TTT GCA GAG AAG TGC TGT AAG (1790)

561 567 570 580
ala asp asp lys glu thr cys phe ala glu glu gln lys lys ala ala ser gln ala ala leu gln leu ter ter
GCT GAC GAT AAC GAG ACC TGC TTT GCC GAC GAC GCT AAA CTT CTT GCT GCA ACT CAA CCT CCC TTA GGC TTA TAA CATCACATTTAAAG (1883)
ter ter
CATCTAGCCCTACCATGAGATAAGAGAAATTGAAGATCAAAGCTTATTCTCTGTTCTCTGGCTGTAAGGCCAACCCCTGCTTAAGAAACATAAATTCTCTTAA (12002)
TCATTTGCCCTTTCTCTGCTTCTCTGGCTGTAAGGCCAACCCCTGCTTAAGAAACATAAATTCTCTTAA..... 20AA (2078)

6. Nucleotide sequence of the cDNA of human serum albumin, said nucleotide sequence is as follows:

5			
10			
15			
20			
25			
30			
35			
40			
45			
50			
55			
60			
65			
70			
75			
80			
85			
90			
95			
100			
105			
110			
115			
120			
125			
130			
135			
140			
145			
150			
155			
160			
165			
170			
175			
180			
185			
190			
195			
200			
205			
210			
215			
220			
225			
230			
235			

1 asp ala his lys ser glu val ala his arg phe lys asp leu ala glu glu asn phe lys
GAT GCA CAC AAG AGT GAC GTT CAT GCT CGC TTT AAA GAT TTG GCA CAA GAA AAT TTC AAA (170)

21 ala leu val ile ala phe ala gln tyr leu gln gln cys pro phe glu asp his val lys leu val asn glu val thr alu phe ala
GCC TTG GTG TTG ATT GCC TTT GCT CAG TAT CTT GAT CAG TGT CCA TTT GAA GAT CAT GTA AAA GAA TTA GTG AAT GAA ACT CAA TTT GCA (260)

51 55 lys thr cys val ala asp glu ser ala glu asn cys asp lys ser leu his thr leu phe gly asp lys leu cys thr val ala thr leu
AAA ACA TGT GTT GCT GAT GAG TCA CCT GCT GAA AAT TGT GAC AAA TCA CTT CAT ACC CTT TTT CGA GAC AAA TTA TGC ACA AAC ACT CTT (350)

81 90 95 arg glu thr tyr gly glu met ala asp cys oys ala lys gln glu pro gly arg asn glu cys phe leu aln his lys asp asp asn pro
CGT GAA ACC TAT GGT GAA ATG CCT GAC TGC TGC TGT GCA AAA CAA CCT CCC GAA CTC TCC TTT GCT TGC TAC AAA AGG TAT AAA TAC TTA TAT (440)

111 120 125 130 135 140 asn leu pro arg leu val arg pro glu val met cys thr ala phe his asp asn glu glu thr phe leu lys tyr leu try
AAC CTC CCC CCA TTG GTG AGA CCA CCT GAT GTC ATG TCC ACT GCT TTT GAT GAC AAA TAT GAA GAG ACA TTT TGA AAA TAC TTA TAT (330)

141 150 155 160 165 170 glu lle ala arg arg his pro tyr phe tyr ala pro glu leu phe ala lys arg tyr lys ala ala phe thr glu cys cys gln
GAA ATT CCC AGA ACA CAT CCT TAC TTT TAT GCC CCC GAA CTC CCT TTG TTT GCT AAA AGG TAT AAA TAC TTA TAT GCA TGT TGC CAA (420)

171 177 180 185 190 195 ala ala asp lys ala ala oys leu leu pro lys leu asp glu leu arg asp glu gln ala ser ser ala lys aln arg leu lys cys
GCT GAT AAA GCT GCT GCC TGC CTC AGC CTC GAT GAA CCT CGC GAT GAA CCT GCT GTC TCT GCT GCA GAA GAG CCT TGC TCC TGT GCA (70)

201 205 210 215 220 225 ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser gln arg phe ala alu glu
CCC AGT CTC CAA AAA TTT GCA GAA AGA CCT TGC GCA GAA AGC TTT CCC AAA GCA GAG CAG AGA TTT GCA GAA GCA (300)

0079739
2083

35 30 25 20 15 10 5

231 240 245 246 250 253 260
val ser lys leu val thr asp leu thr lys val his thr glu cys his gly asp leu glu cys ala asp asp arg ala asp leu
GTT TCC AAG TTA GTC ACA GAT CTT ACC AAA GTC CAC ACC GAA TCC TGC CAT CTC GAA TCA TGT CCT GAT GAC AGC GCG GAC CTT (890)

261 265 270 278 279 280 289 290
ala lys tyr lle cys glu asn gln esp ser lle ser ser lys glu leu cys qlu lys pro leu glu lys ser his cys lle
GCC GAA GTC GAA ATT CAT GAG ATC TCC AGT TCG ATC TCC AGT AAA CTC AAG GAA TGC TGT GAA AAA CCT CTG TTG GAA AAA TGT CAC TGC ATT (980)

291 300 310 316 318 320
ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala esp phe val glu ser lys asp val cys lys amn tyr ala
GCC GAA GTC GAA ATT CAT GAG ATC TCC CCT CCT GAT TAA GCA ACT CCT CAT TAC TCT GTC GTC CTC CTC ACA ATT CCC (1070)

321 330 340 350
glu ala lys esp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val phe glu phe lys amn tyr ala
GAG GCA AAC GAT GTC TTC TGC CCC ATG TTT TTG TAT GAA TAT GCA AGA AGC CAT CCT CAT GAT CTC TCT GTC TGT AAA AAC TAT CCT (1160)

351 360 361 369 370 370
lys thr tyr glu thr leu glu lys cys cys ala ala ala esp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu
AAC ACA TAT GAA ACC ACT CTA GAG AAC TCC TGT CCC ACT GCA GAT CCT CAT GAA TGC TAT GCA TTT AAA CCT CCT (1250)

381 390 392 400 400 410
val glu glu pro gln asn leu lle lys gln asn cys glu leu phe glu aln leu gly glu tyr lys phe gln ASN ala leu leu val arg
GTC GAA AAC CCT CAG AAC ATT ATC AAA CAA ATT TGT GAG CTT TGT GAG CAG TAC AAA TTC CAC AAC ATT GCG CTC TTA TGT CGT (1340)

411 420 430 437 438 440
tyr thr lys val pro gln val ser thr pro thr leu val glu vol ser arg asn leu gly ser lys cys cys lys his
TAC ACC AAC AAA GCA CCC CAA GTC TCA ACT CCA ACT CTT GTA GAC GTC TCA AGA AAC CTA CGA AAA GTC GGC AGC AAA TGT TGT AAA CCT (1430)

441 448 450 460 461 470
pro glu ala lys arg met pro cys ala glu esp tyr leu ser val val phe ser arg pro cys phe ser ala leu glu val asp glu lys thr pro val ser
GAC GCA AAA AGA ATG CCC TGC AGA GAA TCC ACA GAA TCC GTC AAC AGC CCA TCC TGT GTC CAT GAC AAA AGC CCA TAC GTT CCC AAA (1520)

471 476 477 480 490 500
asp arg val val thr lys cys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp glu thr tyr val pro lys
GAG TTT ATT CCT GAA ACA ATT TCC ACA ATT GTC GAC AAC GAG AGA CAA CAA ATT AAC AGC AAA ACT GCA ATT CCT GTC ATT CCC AAA (1610)

501 510 514 520 530
glu phe asn ala glu thr phe thr phe his ala esp lle cys thr leu ser glu lys glu arg aln lle lys aln thr ala leu val
GAG TTT ATT CCT GAA ACA ATT TCC ACA ATT GTC GAC AAC GAG AGA CAA CAA ATT AAC AGC AAA ACT GCA ATT CCT GTC ATT CCC AAA (1700)

0079739
4083

-18-

5

10

15

20

25

30

35

531 540 550 558 559 560
glu leu val lys his pro lys ala thr lys glu gln leu lys ala val met asp asp phe ala ala phe val glu lys cys cys lys
GAG CTC GTC AAA CAC AAC CCC AAG GCA ACA AAA GCA CAA CTC AAA GCA GTC GCT GCT GAT TTC GCT GAT ATG GAT GAT TTT GCA GAG AAG TGC TGC AAC (1790)

561 567 570 580
ala asp asp lys glu thr cys phe ala glu glu gln ala ser gln ala ala leu gln ala leu gln ala leu ter
GCT GAC GAT AAG GAG ACC TGC TTT GCG GAG GAC CGT AAA AAA CTT GCT GCT GCA AGT CAA GCT GCC TTA GCA AAT CAA CATCACATTAAAG (1883)
ter ter

CATCTCAGCCCTACCATGAGATAAGGAAAGAATGAGATCAAAGACCTTATTCACTGTCTTCTTCTGGTTGCTTAACCCACACCTGTCTAAACATAAAATTCTTTAA (2002)
TCATTTGCCCTCTTCTCTGGCTTCAATTAAATAAAAATGGAAAGAATCTAA..... 20AA (2078)

0079739

4083

-19-

7. Nucleotide sequence coding for the prepeptide of human serum albumin, said nucleotide sequence is as follows:

5

10

15

20

25

30

35

GCTTTCTCTCTGTCACCCACACCCCTTGGCAAGATG AAC TGG GAA ATT TCC CTT CTT CTC TTT AGC (30)
-16 p r o Met lys trp val thr phe Ile ser leu leu phe leu thr ser
-10 p r o
-1 p r o -1
ser ala tyr ser arg gly val phe arg arg
TCC CCT TAT TCC ACC CGT GTC TTT CGT CGA

0079739

4083

-20-

8. Nucleotide sequence coding for pro human serum albumin, said nucleotide sequence is as follows:

5	10	15	20	25	30	35
-6 p r o -1						
arg gly val phe arg arg asp ala his lys ser glu val ala his arg phe lys asp leu oly glu asn ohe lys						
AGC GGT CGT TTT CGT CGA CAT GCA CAC AAG AGT GAG TGT CAT CGG TTT AAA GAT TTG GCA GAA AAT TTC AAA						
(170)						
21 ala leu val ala phe ala gln tyr leu gln gln cys pro phe glu asp his val lys leu val asn glu val thr glu phe ala						
GCC TTG CGC TTG ATT GCC TTG CCT GCT CAT GAG TCT CTT CAG CAG TGT CCT GAA AAT TGT GAC TCA TTT GAA GAT CAT GTA ACT CAA TTT GCA (260)						
30 34 38 40 50						
51 53 60 62 66						
lys thr oys val ala asp glu ser ala glu asn oys asp lys ser leu his thr leu phe gly asp lys leu oys thr val ala thr leu						
AAA ACA TGT CTT CCT GAT GAG TCA GCT GAA AAT TGT GAC AAA TCA CTT CAT ACC CCT GAT GAC AAA TGA GTC TTC CAA CAC AAA GAT GAC AAC CCA						
(350)						
81 90 91 95 101						
arg glu thr tyr gly glu met ala asp oys cys ala lys gln glu pro gly arg asn glu oys the leu aln his lys asp asn pro						
CGT GAA ACC TAT CCT GCT GAC TCC TCT GCA AAA CAA CCT GCA AAT GAT GAC AAA TTA TGC ACA GTC TTC CAA CAC AAA GAT GAC CCA						
(460)						
111 120 124 130 140						
asn leu pro arg leu val arg pro glu val met cys thr ala phe his asp asn glu glu thr phe leu lys tyr leu try						
CGT CTC CCC CGA TGT GTC ACT GCT TTT CAT GAC AAT GAA GAG ACA TTT TGC AAA TAC TTA TAT						
(330)						
141 150 160 168 169 170						
glu lle ala arg arg his pro tyr phe tyr ala pro glu leu phe ohe ala lys arg tyr lys ala ala phe thr glu cys qin						
CAA ATT GCC AGA AGA CCT TAC CCT TAC TTT TAT GCT CGG GAA CTC CTT TTC TTT GCT GCT TAT AAA AGG TAT AAA CCT GCT TGT						
(220)						
171 177 180 190 200						
ala ala asp lys ala ala oys leu leu pro lys leu esp glu arg asp glu gly lys ala ser ser ala lys aln arg leu lys cys						
GCT CCT CAT AAA CCT CCC TCC CTC GAT GAA CTT CGC GAT GAA GCG AAG CTC GTC TCT CCC AAG GCT TGC TGT						
(710)						
201 210 220 230						
ala ser leu gln lys ala phe gly gln arg ala trp ala val ala arg leu ser gln arg phe pro lys ala glu ala glu						
GCC ACT CTC CAA AAA TTT GGA GAA CCT TTC AAA GCA TGG CGA GCA CCT CGC AGC CAG AGA TTT CCC AAA GCT GAG						
(360)						

0079739
4083

-21-

35 240 245 246 250 253 255 260
val ser lys leu val thr asp leu thr lys val his thr glu cys cys his gly asp leu leu glu cys ala asp asp arg ala asp leu
GTT TCC AAC TTA GTC ACA GAT CCT ACC AAA GTC CAC ACG GAA TCC TGC CAT CGA GAT CTG CTT GAA TGT GCT CAT GAC AGC GGG GAC CTT (890)

261 265 270 278 279 280 289 290
ala lys tyr lle oys glu aen gln asp ser lle ser lle ser lys glu oys cys glu lys pro leu leu glu lys ser his cys lle
CCC AAG TAT ATC TGT GAA AAT CAA GAT CAC ATG CTC ACT AAA CTC AAG GAA TCC ATC TCC ACT AAA CCT CGT TCA AAA TGT GAA AAC TAC TGC ATT (980)

291 300 310 316 320
ala glu val glu aen asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala
GCC GAA GTG GAA AAT CAT GAC ATG CCT CCT GAC TTG CCT CCT GCT GAT TTT GTC GAA AGT AAC GAT GTT TGT AAA AAC TAC TAT CCT GCC (1070)

321 330 340 340 350
glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp phe val val glu ser lys asp val cys lys ASN
GAG CCA AAG GAT GTC GTC TTC TTG GGC ATG TTT TTG TAT GAA TAT GCA AGA AGC CAT CCT GAT CCT GAT GCA TAC TAC TCT GTC GTC CTC TTC GAT GAA TTT AAA CCT CCT GCC (1160)

351 360 361 369 370 380
lys thr tyr glu thr leu glu lys cys cys ala ala ala asp pro his glu cys tyr ala lys val phe glu phe lys pro leu
AAG ACA TAT GAA ACC ACT CTA GAG AAC TCC TGT CCT CCT GCT GTC TGT TAT GCC AAA GTC AAA GTC TTC GAT GAA TTT AAA CCT CCT CCT (1250)

381 390 392 400 410
val glu glu pro gln asn leu lle lys gln asn cys glu leu phe glu gln leu gly glu tyr lys phe gln asn ala leu leu val arg
GTC GAA GAG CCT CAG AAT TTA ATC AAA CAA AAT TGT GAC CTT GTC CAG CTC CCT GCA GAG TAC AAC TTC CAG AAC GTC CCT TTA GTC CCT (1340)

411 420 430 437 438 440
tyr thr lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu gly lys val gly ser lys cys cys lys his
TAC ACT AAC AAA GTC CCC CAA GTG TCA ACT CCA ACT CTT GTA GAG GTC TCA AGA AAC CTA GGA AAA GTC CGC AGC AAA TGT TGT AAA CAT (1430)

441 448 450 460 461 470
pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val asn gln leu cys val leu his gln lys thr pro val ser
CCT GAA CCA AAA AGA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTC GTC AAC CAG TTA TGT GTC CAT GAC AAA ACG CCA GTC AGT (1520)

471 476 477 480 490 500
asp arg val thr lys cys thr glu ser leu val asn arg pro cys phe ser ala leu glu val asp glu thr tyr val pro lys
GAC AGA GTC ACC AAA TCC TGC ACA GAA TCC TTG GTC MAC AGG CGA TCC CCT CTG CAA GTC GAT GAA ACA TAC CCT CCC AAA (1610)

501 510 514 520 530
glu phe asn ala glu thr phe thr phe his ala asp lle cys thr leu ser glu lys glu arg ala lle lys gln thr ala leu val
GAG TTT AAT CCT GAA ACA ACC TTC CAT ATA TCC ACA CTT TGC AGA AGC AGA CAA ACT GCA CTC GTC CCT GTC (1700)

0079739
4083

-22-

5

10

15

20

25

30

35

531 540 550 558 560
glu leu val lys his lys pro lys ala thr lys glu gln leu lys ala val met asp asp phe ala ala phe val glu lys oys cys lys
GAG CTC GTC AAA CAC AAG CCC AAG GCA ACA AAA GAG GCA GCA AGC TGG CTT CCT GAT GAT ATG GAT GCT TCC GCT GCT TTT GTA GAG AAC TGC TGC AAG (1790)

561 567 570 580 589
ala asp lys glu thr cys phe ala glu glu gln lys lys ala ser gln ala ala leu gln leu ter
GCT GAC GAT AAG GAG ACC TGC TTT GCC GAG GAG GGT AAA AAA CCT GTT GCT GCA AGT CAA GCT CCC TTA GGC TAA CATCACATTAAAG (1881)
ter ter ter ter ter
CATCTCAGCTTACCATGAGAATAGAGAAAGAAAATGAAGATCAAAGCTTATTCTCATCTGTTTCTCTGTTGCTGTTAAACCCCTGCTAAAAAACATAAATTCTTTAA (2078)
TCATTTGCCCTCTTCTCTGCTGCTCAATTAAAAAATGGAAAGAATCTAA.... 20AA (2078)

0079739
4083

-23-

9. Nucleotide sequence coding for the pre pro human serum albumin, said nucleotide sequence is as follows:

35	30	25	20	15	10	5
					-10	-10
					p r c	
						Met lys trp val thr ile ser leu leu phe leu the ser
						ATG AAC TGC GTA ACC TTT ATT TCC CTT CTC TTT AGC (130)
						TCG GCT TAT TCC AGC GGT GTG TTT CGT CGA GAT CGC GAC AGT GAG (170)
		21	30	34	40	50
						ala leu val ile ala phe ala gln tyr leu gln gln cys pro phe glu asp his val lys leu val asn glu val thr ala phe ala
						GCC TTG GTC TTG ATT GCC TTT CCT CAG TGT CGA AAA GAT CAT GCA TAA TCA CTT CAT ACC CTT CAT GAC AAA TGT GAC GAA ACT CTT GCA ACT CTT GCA ACT CTT (260)
		51	53	60	62	70
						lys thr cys val ala asp glu ser als glu asn cys asp lys ser leu his thr leu phe gly asp lys leu cys thr val ala thr leu
						AAA ACA TGT GTT GCT GAT GAG TCA GCT AAC AAT TGT GAC AAA TCA ATT GCA ATG CCT GCA ACC TAT GGT CAA ATG GAA CTC CCC (350)
		81	90	91	100	101
						arg glu thr tyr gln glu met ala asp cys als lys gln glu pro gly arg asn glu asn glu thr phe leu gln his lys asd asn pro
						CGT GAA ACC TAT GGT GAA ATG CCT GAC TGC TGT CGA AAA CAA CCT GGG AGA AAT GAA TCC TTC TGC CAA CAC AAA GAT GAC AAC CCA (440)
		111	120	124	130	140
						asn leu pro arg leu val arg pro glu val met cys thr als phe his asd asn glu glu thr phe leu lys lys tyr leu try
						AAC CTC CCC CCA TTC GTC AGC ACA CCA CTC GGT GAT GTC ATG TGC ACT GCT TTT CAT GAC AAT GAA GAG ACA TTT TTG AAA AAA TAC TTA TAT (330)
		141	150	160	168	169 170
						glu ile als arg his pro tyr phe tyr als pro glu leu phe the als lys arg tyr lys als ala phe thr alu cys cys qln
						CAA ATT CCC AGA CAT CCT TAC TAC TAT GTC CGC CGA CTC CTT TTC ATT CCT AAA AGC TAT AAA GCT CCT TTT ACA GAA TGT TGC CAA (620)
		171	177	180	190	200
						ala als asp lys als als cys leu leu pro lys leu arg asp glu gly lys als ser ser als lys als ala leu lys cys
						GCT GAT AAA GGT GCC TGC TTC CGA AGG CTC GAT GAA CCT CGC GAT GAA GCG AAG CCT TCG TCT GCC AAA CAG AGA CTC AAG TGT (710)
		201	210	220	230	
						ala ser leu gln lys phe gly glu arg als phe lys als val als arg leu ser gln arg als phe als glu
						GCC AGT CTC CAA AAA TTT GGA GAA GCA GCT TTC AAA CCT CCC CTG ACC CAG AGA TAA GGT GAC TTT GCA GAA (330)

0079739

4083

-24-

0079739

4083

-25-

5

10

15

20

25

30

35

531 540 550 558 559 560
glu leu val lys his lys pro lys ala thr lys glu gln leu lys ala val met asp asp phe als als als phe val glu lys cys cys cys lys
CAG CTC GTC AAA CAC ACG CCC AAG GCA ACA AAA GCA CAA CTG AAA GCT GTT ATG GAT CAT TTC CCT GCT GTC TCC GCG AAG TCA GAG AAC (1790)

561 567 570 580
ala asp lys glu thr cys phe als glu glu qly lys leu val als als ser gln als als leu ala leu ter ter
GCT GAC GAT AGC GAG ACC TGC TTT GGC GAG CAC GGT AAA AGC AGT CAA AGT GCA AGT CAA GCT GCC TTA TAA CATCHCATTTAAC (1683)

ter ter
CATCTAGGCTACCATGAGATAAGCAGAAATGAGCTAAAGCTTATTCTCTGTCTTCTTTGGTGCTGGTCAACCTGTCTTAAACATAAATTCTCTAA (2002)

TCAATTTGCCCTCTTTCTCTGCTCTAAATTAAAATGGAAAACATCTAA..... 20AA (2078)

10. A nucleotide sequence according to any of claims 6 to 9, in essentially pure form.
11. A DNA transfer vector comprising a nucleotide sequence as defined in claim 5.
- 5 12. A DNA transfer vector according to claim 11, transferred to and replicated in a micro-organism.
13. A DNA transfer vector according to claim 12, which is a plasmid.
14. A DNA transfer vector according to claim 13, 10 wherein the plasmid is pBR322 or YEp6.
15. A process for preparing human serum albumin, which comprises culturing a micro-organism according to claim 5.
16. A DNA transfer vector according to any of 15 claims 12 to 14, or a process according to claim 15, wherein the micro-organism is a bacterium or yeast.
17. A vector or process according to claim 16, wherein the bacterium or yeast is E. coli or Saccharomyces cerevisiae.

0079739

Restriction Endonuclease Map of Human Serum Albumin cDNA Clones

